



Explainable liver tumor delineation in surgical specimens using hyperspectral imaging and deep learning

YATING ZHANG,¹ SI YU,² XUEYU ZHU,³ XUEFEI NING,¹ WEI LIU,⁴
CHUTING WANG,² XIAOHU LIU,^{1,5} DING ZHAO,¹ YONGCHANG
ZHENG,^{2,6} AND JIE BAO^{1,7}

¹Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

²Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100730, China

³Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA

⁴Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China

⁵Currently with the School of Biomedical Engineering, School of Ophthalmology & Optometry, Wenzhou Medical University, Xueyuan Road 270, Wenzhou 325027, China

⁶zhengyongchang@pumch.cn

⁷bao@tsinghua.edu.cn

Abstract: Surgical removal is the primary treatment for liver cancer, but frequent recurrence caused by residual malignant tissue remains an important challenge, as recurrence leads to high mortality. It is unreliable to distinguish tumors from normal tissues merely under visual inspection. Hyperspectral imaging (HSI) has been proved to be a promising technology for intra-operative use by capturing the spatial and spectral information of tissue in a fast, non-contact and label-free manner. In this work, we investigated the feasibility of HSI for liver tumor delineation on surgical specimens using a multi-task U-Net framework. Measurements are performed on 19 patients and a dataset of 36 specimens was collected with corresponding pathological results serving as the ground truth. The developed framework can achieve an overall sensitivity of 94.48% and a specificity of 87.22%, outperforming the baseline SVM method by a large margin. In particular, we propose to add explanations on the well-trained model from the spatial and spectral dimensions to show the contribution of pixels and spectral channels explicitly. On that basis, a novel saliency-weighted channel selection method is further proposed to select a small subset of 5 spectral channels which provide essentially as much information as using all 224 channels. According to the dominant channels, the absorption difference of hemoglobin and bile content in the normal and malignant tissues seems to be promising markers that could be further exploited.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Liver cancer is the sixth most commonly diagnosed cancer and the fourth leading cause of cancer death in the world, with approximately 841,000 new cases and 782,000 deaths each year [1]. Surgical removal is the primary treatment, but studies have shown that 30% to 50% of patients suffer from the recurrence, which greatly affects the long-term survival rate of patients [2]. Malignant tissues left during the surgery are one of the closely related factors affecting the recurrence of liver cancer. However, the golden standard assessment of specimen can only be obtained through the pathology reports after more than five to seven days, which is not able to provide timely feedback to the surgeon during the surgery. In this regard, intra-operative visual inspection is necessary. Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) mainly cover over 90% of primary liver cancer. Currently, clinical intraoperative imaging

modalities have been restricted to a certain extent. Intraoperative ultrasound (IOUS) is not suitable to perform the real-time guidance due to its contact imaging characteristics [3]. Compared with IOUS, fluorescence-guided surgery (e.g., Indocyanine green (ICG)-fluorescence imaging), as an optical imaging method, performs well, especially on superficial tumors in a real-time and high-sensitive way. But the use of contrast agent brings side effects or a relatively high false-positive rate of 40% [4,5]. Therefore, there is an urgent need to develop a fast, sensitive and label-free imaging modality.

Recently, hyperspectral imaging (HSI), as a sensitive optical imaging modality, becomes one of the promising candidates for its label-free imaging characteristics. Most traditional optical imaging modalities can only capture one channel (gray) or three-channel (color) images of biological samples, while HSI captures an extra dimension of image data, i.e., the spectral dimension (Fig. 1). HSI is a fast-evolving optical modality that combines high spectral and spatial resolution in one setup, which has been proved to be promising for medical diagnosis [6]. The spatial information describes the size and morphology of tissue and the spectral information acts as fingerprints of tissue components. Based on this advantage, HSI can greatly help surgeons expand their vision into invisible wavelengths, which enables fast, sensitive, and label-free tumor delineation.

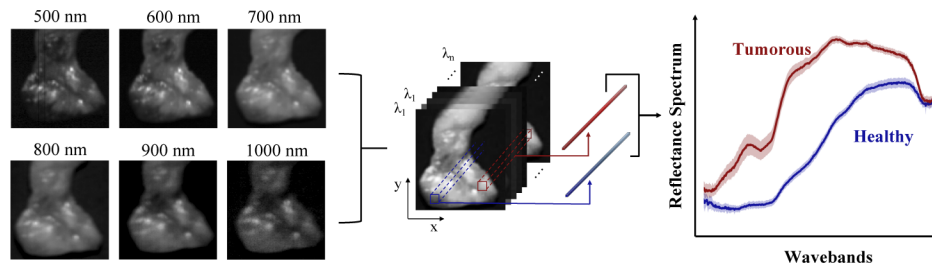


Fig. 1. Illustration of the spatial and spectral characteristics of a hyperspectral image. Left: visualization of an example specimen under wavelength 500 nm-1000 nm. Middle: a hyperspectral image data cube consists of two spatial dimensions (x , y) and one spectral dimension (λ), which can be seen as spectrally resolved images or spatially resolved spectra. Two bounding boxes capture local pixels of primary liver tumor tissue (red) and normal liver tissue (blue), respectively. All the spectra in these selected local areas (shadow area) and their average spectral curves across the channel (solid lines) are displayed on the Right, illustrating the spectral differences between these two different tissues.

The biological explanation of spectra in hyperspectral images can be explained as [7]. Due to the existence of biological tissue structure and chromophore, light undergoes a series of scattering and absorption when transmitted in the tissue. Tissue absorption is a function of molecular composition. When the photon's energy matches the gap between the internal energy states, the molecule will absorb the photon. Since molecules have different energy levels, the photon energy they can absorb varies accordingly, which serves as a spectral fingerprint for molecular diagnosis. Besides, scattering is related to the structure and number of organelles in the cell that have a refractive index difference from the surrounding medium [8]. In the process of tumor development, as the composition of the tissue changes, its pathological characteristics, light absorption and scattering characteristics also change correspondingly. In this way, the spectral characteristics of tissue captured by HSI can be converted into quantitative diagnostic information.

Previous studies reported some promising results of HSI in tumor surgeries, e.g., glioblastoma tumor surgery [9–11], head and neck cancer surgery [12,13], and breast-conserving surgery [14,15], which basically focus on fast and sensitive tumor detection and margin assessment. At present, little effort can be found on the application of the label-free HSI in liver tumor surgery.

In this paper, for the precise intraoperative delineation of primary liver tumors, we collected a clinical hyperspectral image dataset of fresh ex-vivo liver tissues. Using the dataset, a U-Net framework was developed to automatically segment tumors on surgical tissues in order to improve the performance of automatic delineation to approach the post-operative gold standard.

Owing to improvements in image analysis via deep learning, much progress has been made in automated medical diagnosis on hyperspectral images [15,16]. However, how these models generate predictions based on input data still remains a black box, particular in safety-critical applications, such as medical diagnosis, human-interpretable explanations are essential for the results [17]. Guided by visual appeal, existing studies [18–20] that consider interpretability mainly focus on extracting salient spatial regions of the input image. These regions are relevant for the prediction of a learned model. In medical HSI applications, model interpretability has not been extensively studied.

In addition to spatial saliency, the spectral interpretability of hyperspectral images is also a question worth exploring. Since the spectral information is related to tissue composition, it can provide us with some biomedical insights from the model's decision. In this paper, we employ a saliency map-based method to interpret the outputs of our tumor delineation model. Upon that, a very small subset of channels with representative diagnostic information is selected through our proposed method for more specific spectral indications. The benefits of selecting a small subset of spectra are two-fold: i) The more detailed channels may reveal the biological value of specific wavelengths more explicitly; ii) Reducing the spectral dimension can effectively improve the hardware/software codesign by reducing system filter integration, data collection time and model complexity. Thus, our method demonstrates the potential of a low-latency HSI system design for real-time intra-operative use in the future.

The contributions of this paper are summarized as follows.

- To the best of our knowledge, this is the first work evaluating the feasibility of HSI on primary liver tumor surgery. We collect a clinical hyperspectral image dataset of fresh ex-vivo liver tissues on 19 patients. With pathological annotation as ground truth, we develop a U-Net deep learning framework to automatically delineate the tumor margin on fresh specimens.
- We propose to add post-hoc explanations to this framework to show the interpretability of the model's decision. A saliency map-based method is applied to the well-trained model to extract salient pixels and spectral channels from the hyperspectral image data. From the spectral interpretability, we propose a novel saliency-weighted method to select a very small but representative subset of spectral channels for the specific task.
- Experimental results demonstrate the effectiveness of our framework. The tumor delineation network achieves an overall accuracy of 93.16% and a sensitivity of 94.48%. In addition, we show that only using 5 spectral channels is sufficient to reproduce the result of using 224 channels within a small sensitivity degradation of 2%, and even outperform the former in accuracy and specificity.
- Notably, our investigation reveals the correlation of salient spectral channels with differences in absorption between the normal and malignant liver tissues, indicating that the absorption characteristics of hemoglobin and bile at 420nm and 605 nm might be promising indicators to be further exploited.

2. Methods

2.1. Study design

The ex-vivo study includes patients who underwent primary liver cancer surgery at Peking Union Medical College (PUMC) Hospital (Beijing, China). For the experiment design, basically

following [13], we collected specimens during the surgery and the corresponding pathological result could be obtained in five to seven days. To evaluate the capability of HSI to detect and delineate tumorous tissues, specimens containing both tumor and adjacent healthy liver tissues are included in this study. The overview workflow of dataset collection is shown in Fig. 2. Tissue slices are sampled without affecting the normal clinical process.

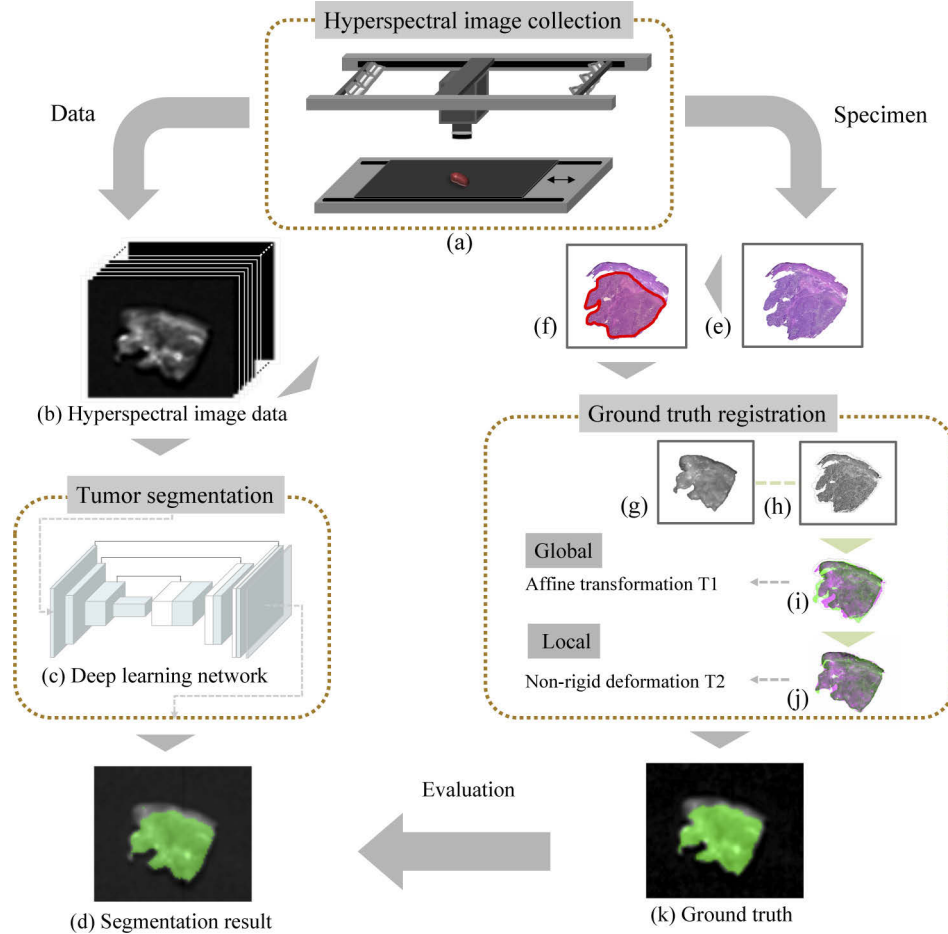


Fig. 2. Overview of the dataset collection and result evaluation process. Surgical specimens are measured using an HSI setup (a). The data obtained from the HSI measurement is in the form of an image data cube (b), which is then processed by the automatic segmentation framework (c) to obtain a fast rendering result (d). The measured specimens are processed into H&E-stained sections (e) according to the standard pathological process, and then the tumor margin (f) is outlined by professional pathologists under the microscope. To obtain a ground truth annotated on the hyperspectral macro images, gray scale images of the RGB image (g) synthesized from the hyperspectral datacube and the pathological section image (h) were used for the registration. The registration follows a pipeline of global affine alignment (i) and a local deformation refinement (j). Accordingly, the ground truth on pathological sections can be mapped onto hyperspectral macro images (k), serving as supervision information for the auto-delineation network.

After resection, we carry out the hyperspectral image measurement of the fresh specimens in less than 15 minutes (Fig. 2(a)). During the measurement, a light absorbing black-out material is placed under the tissue so as to avoid the interference of background stray light and followed

by the standard pathological process. The measured specimens are then processed through fixation, slicing, hematoxylin and eosin (H&E) staining and then analyzed under microscopy by professional pathologists. The gold standard results on pathological sections outlined by pathologists are later mapped to the macroscopic spectral images, serving as the ground truth for algorithm evaluation (Fig. 2(k)). So far, all hyperspectral images of specimens are paired to their corresponding pathological ground truth, thereby constructing a dataset for automatic tumor delineation. Upon that, we developed a deep learning network for tumor segmentation (Fig. 2(c)) and added a post-hoc saliency explanation on hyperspectral images. The performance of tumor delineation was evaluated by comparing its results with the pathological gold standard, proving the effectiveness of the HSI combined with deep learning models.

2.2. System specifications

Hyperspectral images in the dataset were acquired using a pushbroom hyperspectral camera (SPECIM FX10, Spectral Imaging Ltd., Finland) that captures light in the visible and near-infrared (400-1,000 nm spectral coverage, 224 wavelength bands) with a CMOS sensor of 1024×1 pixels' resolution. To obtain a full view of the target specimen, samples were placed upon a translational platform and then scanned underneath the camera. The captured images have a size of 224 wavelengths \times 1024 pixels \times a varying number of scanned lines. Six halogen light sources (DECOSTAR 51 ALU, OSRAM) were used for illumination from both sides, under an angle of 45 degrees. Fig. 2(a) is a schematic diagram of the HSI setup.

During every measurement, a dark current and a white reference image are also captured for data calibration. This preprocessing step is to convert raw hyperspectral images into normalized diffuse reflectance so as to correct for spectral non-uniformity of the illumination and the influence caused by dark current [14]. The dark reference image is taken by closing the shutter of the camera. The white reference image is taken using a Spectralon target (SRT-50-050 Reflectance Target, ACAL Bfi Nordic AB, Uppsala) with known reflectance, providing the correction of the uneven illumination across the field of view. The linear behavior of the visual camera allows for a simple calibration of data by using

$$X_{\text{refl}} = \frac{X - R_{\text{dark}}}{R_{\text{white}} - R_{\text{dark}}}, \quad (1)$$

where X_{refl} denotes the calibrated spectrum, X is the original measured spectrum. R_{dark} and R_{white} represent the dark and white reference, respectively.

2.3. Ground-truth annotation

After HSI measurement, specimens are processed into pathological sections through fixation, slicing and staining, etc. During the section preparation phase, tissue undergoes deformations, thus the pathological micro images are different from hyperspectral macro images in both shape and morphology. To obtain a ground truth on the hyperspectral image, registration is required to map the tumor margin from the pathological section to the gross specimen. According to the basic pipeline in [21–23], registration is carried out in two steps, namely, global boundary alignment and local deformation refinement. Before the registration, we convert both images to grayscale and add Gaussian smoothing to reduce the impact of morphological differences. First, a global affine transformation is applied using mutual information as a metric to measure the similarity between multi-modal images. After that, the boundaries of two images are aligned through rigid transformations (Fig. 2(i)). An obtained global transformation matrix T_1 is then applied to the pathological image. In step two, structural landmarks are selected in both pathological images and hyperspectral images as the control points for the subsequent detail fine-tuning. Then, a B-spline free-form deformation (FFD) non-rigid registration is applied to model the local deformation in the tissue [24]. The spline function-based registration method mainly uses control

points and spline functions to describe the nonlinear geometric transformation domain, which can effectively fit the elastic deformation of the object. We apply the deformation registration to the images which have a size within 320×320 pixels with an image sampling rate of 1.34 mm. A control point spacing of 32 mm or 20 mm yields improved correlation compared to the global transformation. The registration quality was measured by the sum of squared intensity differences. A strategy of Limited-memory BFGS (L-BFGS) [25] is applied to find the optimal deformed grid T2. Finally, the deformed grid T2 is applied to the transformed image from step 1, thereby the tumor margin is globally and locally registered (Fig. 2(j)). Accordingly, the outlined tumor margin from the microscopic image is mapped onto the optical image of the gross specimen, i.e. a hyperspectral image data set of specimens with pathological ground truth is obtained. In addition, every specimen is labeled with an image-level diagnostic class of the absence or presence of the primary liver tumor. All registrations are implemented using MATLAB (R2020a).

2.4. Tumor delineation in surgical specimens

Model design We carry out the tumor margin auto-delineation in the specimens following a multi-task strategy of segmentation and classification. Since the U-Net architecture [26] achieves outstanding results in medical segmentation tasks with a small data volume, we developed the U-Net architecture on HSI image dataset following its encoder-decoder scheme. The encoding network performs down-sampling which extracts features from both spatial and spectral dimensions with multiple encoder blocks. The decoding network upsamples the feature map to input image size using multiple decoder blocks. Skip connections between the encoder and the decoder provide finer information to the decoding network, thus improving the segmentation result to a large extent.

The multi-task network is implemented by adding a classification branch to the U-Net framework for the subsequent explanation of the decision-making process. Taking the hyperspectral images patches as the input, one branch outputs the pixel-level segmentation mask of the liver tumor while the other outputs an image-level classification result, indicating the absence or presence of the tumor. The network follows a fully-convolutional encoder-decoder architecture. In the encoding branch, all the convolution blocks are followed by batch normalization and ReLU activation. The detailed architecture is shown in Fig. 3. The model is trained with a multi-task loss $L = L_{seg} + L_{cls}$, where the categorical cross entropy loss L_{seg} with the softmax function can be described as:

$$L_{seg}(y, \hat{y}) = -\frac{1}{N} \sum_i^N \sum_k^K [\mathbf{1}\{y_i = k\} \cdot \log p_{model}(y_i = k)]. \quad (2)$$

Here, i is the sample number, k is the category number, and $K = 3$. The term $\mathbf{1}\{y_i = k\}$ is the indicator function of sample i belonging to class k , and $p_{model}(y_i = k)$ is the model's prediction \hat{y} . In the case of classification, $K = 2$ and the binary cross entropy loss L_{cls} can be specified:

$$L_{cls}(y, \hat{y}) = -\frac{1}{N} \sum_i^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (3)$$

Data preprocessing Hyperspectral images are divided into 64×64 small patches as the model input, and a leave-one-patient-out cross-validation is applied to evaluate the model's performance. The number of tumor pixels is fewer than that of other tissues, which causes a serious imbalance among classes. Resampling patches of different classes with an equal ratio can be applied to effectively solve the issue [27]. In order to improve the generalizability of the model, data is augmented from both spatial and spectral dimensions during the training [28]. Standard geometric transformations are applied on spatial images, such as random horizontal and vertical flipping, rotations, scales, and translations. In addition, we add radiation noise on every spectrum

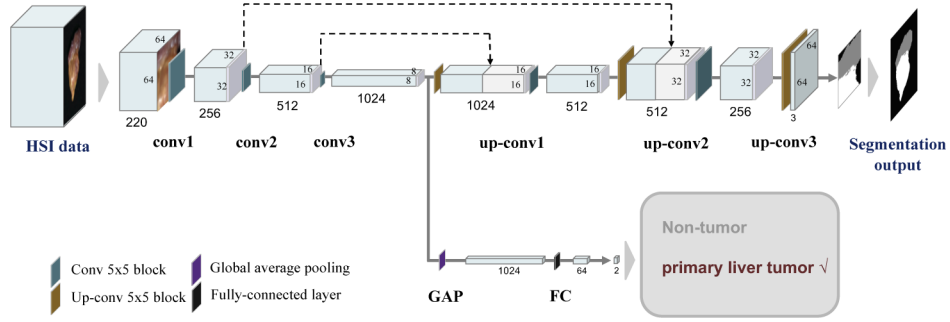


Fig. 3. The developed U-Net framework for hyperspectral image segmentation and a classification branch for post-hoc explanation.

to improve the invariance of the model to spectral noise. In the experiment, a multiplicative noise generated from a uniform distribution ranging from 0.9 to 1.1 and an additive noise drew from a normal distribution are applied on every single spectrum as the spectral augmentation.

2.5. Saliency in hyperspectral images

The performance of the tumor delineation model can be evaluated by the gold standard, i.e., the annotation of pathological sections, indicating the reliability of the segmentation results. However, the annotation can inevitably have subtle uncertainties caused by manual delineation and auto-registration. For a reliable medical decision, we added post-hoc explanations to further evaluate the result. A rather simple but intuitive saliency map strategy is proposed to look into the model and further improve the overall design of the system (see Fig. 4). Saliency maps have been used as a popular visualization technique to detect how and why a deep learning neural network makes certain predictions [19]. We applied it to show the image-specific saliency of a given class, assessing the result of segmentation.

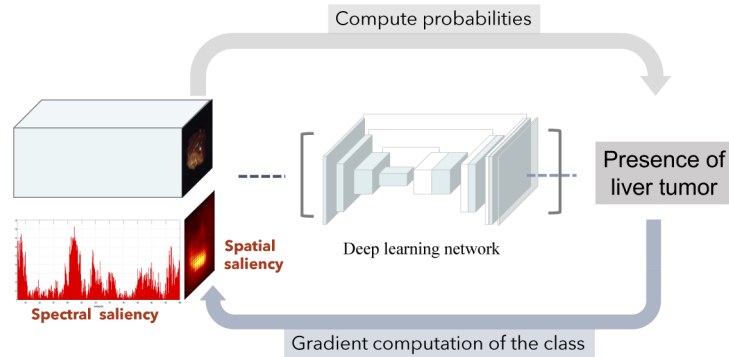


Fig. 4. The post-hoc saliency explanation of the model's decision on a hyperspectral image.

Compared with previous work, we focused on providing saliency explanations from both spatial pixels and spectral channels. Through one single back-propagation, the derivative ω from a specific predicted category to the input image can be obtained from the well-trained model with the index $h(i, j, c)$. Here, (i, j) indicates the spatial arrangement of elements in ω , while c indicates spectral channels. From the aspect of spatial saliency, the map M is computed as

$$M_{i,j} = \max_c |\omega_{h(i,j,c)}|, \quad (4)$$

in the case of a multi-channel image, which is the same as a three-colored image. Inspired by the spatial saliency, the spectral saliency scores \mathbf{s} can be accordingly specified as

$$\mathbf{s}_c = \sum_i \sum_j |\omega_{h(i,j,c)}|, \quad (5)$$

through summing all contributions of spatial pixels under one specific channel, which represents the contribution of different spectral channels to the classification result.

2.6. Spectral channel selection

Since adjacent spectral channels are highly correlated, it is desirable to eliminate channels that may not carry discriminative information in this task. Previous work reported that a reduction of spectral channels could actually reduce the complexity of high-dimensional data like hyperspectral images and then improve the results [29,30]. However, strategies such as unsupervised correlation analysis [31,32] in the preprocessing step lack task adaptation, and the method of supervised training requires labeled data [33] to select discriminative channels. Here, we propose a novel method to find a discriminative subset by introducing the channel saliency (see Section 2.5).

Let $C = \{\mathbf{c}_i\}_{i=1}^n$, $\mathbf{c}_i \in \mathbb{R}^m$ be a ground spectrum set, where \mathbf{c}_i denotes the normalized spectral feature vector of a channel i . Here, n is the maximum number of spectral channels and m represents the length of the data sets. Specifically, we assign the saliency scores $\{\mathbf{s}_i\}_{i=1}^n$, $\mathbf{s}_i \in \mathbb{R}$ as weight to the spectral feature vector as a task-specific prior for the channel selection process. Consequently, the saliency-weighted spectral features can be obtained as

$$C' = \{\mathbf{c}_i \mathbf{s}_i^\sigma\}, \mathbf{c}_i \in \mathbb{R}^m, \mathbf{s}_i \in \mathbb{R}, \quad (6)$$

where σ is an optional tuning factor that controls the influence of the saliency score.

Thus, the task-specific subset selection problem can be described as: given an $m \times n$ saliency-weighted data set C' with $m \geq n$, the objective is to select the r significant channels out of the n channels or to select an $m \times r$ subset $C_1 (r < n)$ of C' , which contains the most salient information in this task. Channel selection is carried out by deterministic and randomized strategies respectively, that is, QR factorization with column pivoting (QRcp) and determinantal point processes (DPP) sampling. Note that in this paper, we mainly emphasize the idea of including the saliency scores into the subset selection process and specify the modifications added in these methods. Mathematical notions and implementation details can be found in [34–37].

QRcp A column-pivoted QR factorization method [35,38] is adopted as a deterministic subset selection method with the rank r . It is applied to capture the most salient channels by choosing those discriminative features that show minimal similarities among them in an orthogonal sense. Here in this task, columns in C' denote the L_1 -norm saliency-weighted spectral features in each channel. The column pivoting option allows us to detect dependencies among the channels of the data set. The column order is determined in a greedy fashion. The first column of C' is selected according to the largest Euclidean norm weighted by its saliency, which means a more salient channel for this task with more independent characteristics with adjacent channels is more likely to be chosen as the first column of C_1 . Successively, the column having a maximum orthogonal component to the first selection is selected, and so on.

DPP sampling As an elegant probabilistic model, the DPP model [36,37] is adopted as a randomized method to find an r -element subset. According to the method, the probability assigned by a DPP to a subset C_1 is proportional to the determinant of the covariance kernel matrix L , that is

$$P(C_1 \subseteq C) \propto \det(L_{C_1}), \quad (7)$$

according to the geometric interpretation of determinant, i.e., the square of the volume spanned by the spectral feature vectors \mathbf{c}_i . Therefore, as the magnitude of the channel's feature vector

increases, so do the probabilities of sets containing that channel. Accordingly, highly correlated spectral channels are less likely to co-occur in the subset, so the diversity can be encoded during the sampling process [31]. Specifically, here the kernel L is modified with the consideration of the L1-norm saliency

$$L = (C')^T C' = (\mathbf{c}_i \mathbf{s}_i^\sigma)^T (\mathbf{c}_j \mathbf{s}_j^\sigma), i, j \in \{1, 2, \dots, N\}. \quad (8)$$

Finally, the r -element subset is determined with the maximum probability.

3. Results

In this section, we first introduce the basic settings of the following experiment. Then, we present the tumor delineation result on the constructed dataset and compare its performance with the baseline method. A saliency map-based post-hoc explanation and channel selection method was employed on the well-trained model to explore the human-interpretable difference between tissues. The experiments are conducted under an inter-patient framework to evaluate the model's transferability on different individuals. We also discuss the result of intra-patient settings with the same network and emphasize the influence of tumor heterogeneity.

3.1. Experimental setup

Dataset description In this work, 19 patients and a total of 36 surgical specimens were included in this study, with a ratio of 12 HCC cases and 7 ICC cases. As described in Section 2.1, we constructed a hyperspectral image dataset for primary liver tumor tissues with pathological annotations serving as the ground truth. Note that a part of the specimens were taken from different locations of the same patient to evaluate the intra-patient performance. The characteristics of patient and specimens are shown in Table 1.

Table 1. Patients characteristics

Patient ID	Number of Specimen	Age	Gender	Histological Type	Differentiated
1	2	62	M	HCC	Moderately
2	1	82	M	HCC	Moderately
3	2	76	F	ICC	Moderately - Poorly
4	2	52	M	HCC	Moderately - Poorly
5	2	43	F	ICC	Moderately - Poorly
6	1	67	M	ICC	Poorly
7	3	41	M	HCC	Moderately
8	1	57	M	HCC	Moderately
9	3	50	M	HCC	Poorly
10	3	63	M	ICC	Moderately - Poorly
11	3	57	F	ICC	Moderately - Poorly
12	1	62	M	HCC	Moderately
13	1	80	M	ICC	Moderately
14	2	65	M	HCC	Poorly
15	1	49	F	ICC	Moderately
16	1	80	M	HCC	Well - Moderately
17	4	47	M	HCC	Moderately
18	2	83	M	HCC	Well - Moderately
19	1	72	F	HCC	Moderately

With annotation on hyperspectral images, each pixel/spectrum was labeled into a class: primary liver tumor, normal liver tissue or background. All the spectra of different tissues and their average spectra included in this dataset are given in Fig. 5. As we can see, although the spectrum has a certain offset in amplitude, the shape of the spectrum from the same type of tissue is similar. In our dataset, the spectra of HCC and ICC show high similarity in spectral characteristics, thus labeled into the same class, i.e., primary liver tumor. They all show obviously different characteristics from normal tissues.

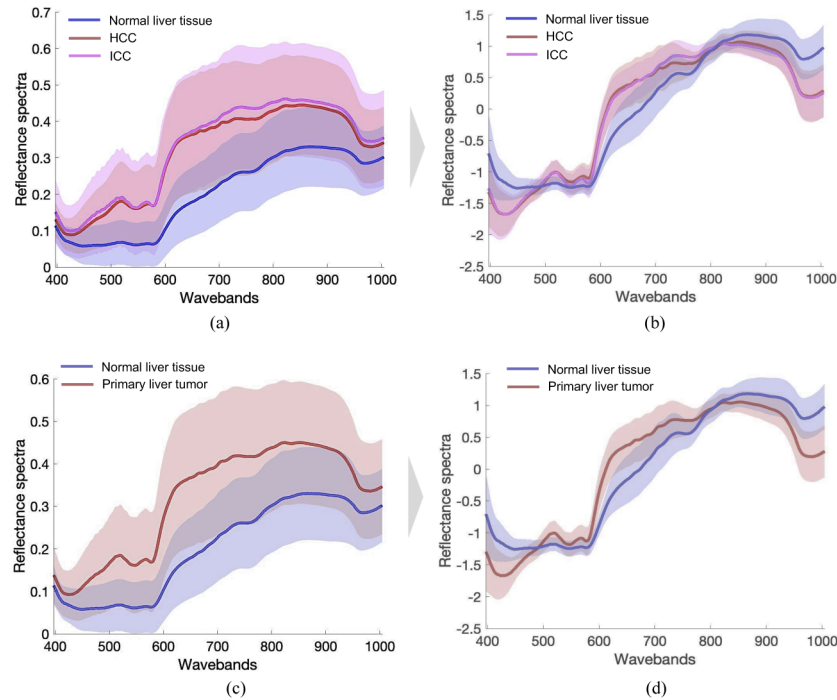


Fig. 5. Spectral reflectance curves of all data in the dataset are depicted (shadow) with the average spectra (solid curve) of (a) normal liver tissue, HCC and ICC; (c) normal liver tissue and primary liver tumor. The normalized spectra were shown in (b) and (d), respectively.

Network settings On each specimen image, we randomly sampled 60 patches of tumor, normal tissue and background with an equal ratio. Each patch is assigned a label indicating the presence or absence of liver tumors. The experiment was carried out in a leave-one-patient-out cross-validation setting. In each cross-validation, a data set of around 6030 patches of size $64 \times 64 \times 224$ from 18 patients were used as the training set and around 120 patches from 1 patient was left for the validation. The model is optimized using stochastic gradient descent with a base learning rate of 0.001, decayed by a factor of 2 after 30 epochs in a total of 100 epochs. All experiments were conducted on a NVIDIA DevBox equipped with four TITAN GPUs, and all deep learning models were implemented using PyTorch.

Baseline methods A support vector machine (SVM) based classification model was developed acting as the baseline method. Rather than splitting patches, the pixel-based classification method considers a hyperspectral image as a library of spectra to be distinguished from different signatures. The SVM model separates different tissues by determining an optimal hyperplane that maximizes the margin between different classes in higher-dimensional space. In this regard, SVM works on the spectrum of every single pixel, which means the texture information is neglected.

With the same patch setting in the U-Net framework, the multi-class classification was carried out using SVM through a one-against-all strategy [39].

Performance metrics The performance was measured by metrics of the accuracy, sensitivity, specificity and a Dice similarity coefficient (DSC) according to the definition of true negative (TN), false negative (FN), true positive (TP) and false positive (FP). The metrics are computed as

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad \text{Specificity} = \frac{TN}{TN + FP},$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \quad \text{DSC} = \frac{2TP}{2TP + FP + FN}.$$

3.2. Experimental results

3.2.1. Tumor delineation

Inter-patient study The results of tumor delineation on surgical specimens ($N = 36$) show the effectiveness of using U-Net architecture compared with the baseline SVM method. Figure 6 presents the performance of the two methods in terms of accuracy, sensitivity, specificity, and DSC. The developed U-Net framework outperforms SVM on this task by a substantial margin, reaching an overall accuracy of $93.06 \pm 4.60\%$, a sensitivity of $93.75 \pm 6.24\%$, a specificity of $87.22 \pm 11.23\%$, and an DSC of $92.91 \pm 6.93\%$ with lower standard deviation. Figure 7(d) and (e) give the visual delineation results of both methods to discriminate the malignant liver tumor from the normal liver tissue. It can be seen that SVM shows more noise in the results associated with the margin where normal and malignant tissues intersect. Moreover, SVM performs worse on ICC specimens (see Fig. 7 S1 and S6) than on HCC samples, which is manifested in the appearance of some salt and pepper noises. This greatly degrades its sensitivity performance. It is also worth mentioning that under the same experimental settings, U-Net is also superior to SVM in terms of time performance, saving training time by more than 3 times. Under the network settings in Section 3.1, it takes the U-Net architecture 4-5 hours to perform cross-validation training for 19 patients. When evaluating on the forward model, it takes 0.71s on average to segment and classify one sample image.

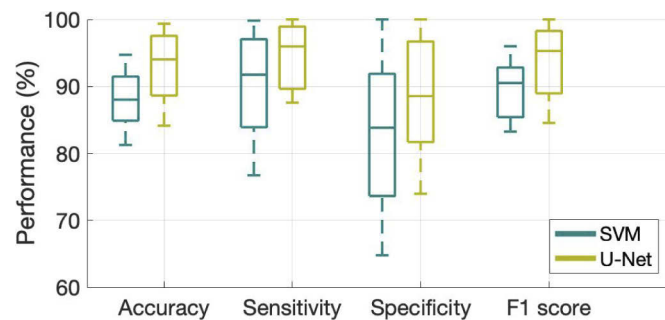


Fig. 6. The boxplots depict the model performance of the developed U-Net model and the baseline SVM model in different metrics.

Ablation study Ablation study was performed over a number of facets of the model in order to better understand their effects on the performance. We mainly ablated the influence caused by the size of the receptive field, spectral data augmentation, spectral normalization, number of patches sampled in every image, class reweight on patches and batch size, etc (see Table 2).

Especially, there are three factors worth mentioning: First, the size of the receptive field is not negligible. Three commonly used kernel sizes ranged from 1, 3 to 5 were tested (E0-E2). The results suggest that the increase of the receptive field obviously helps improve the model

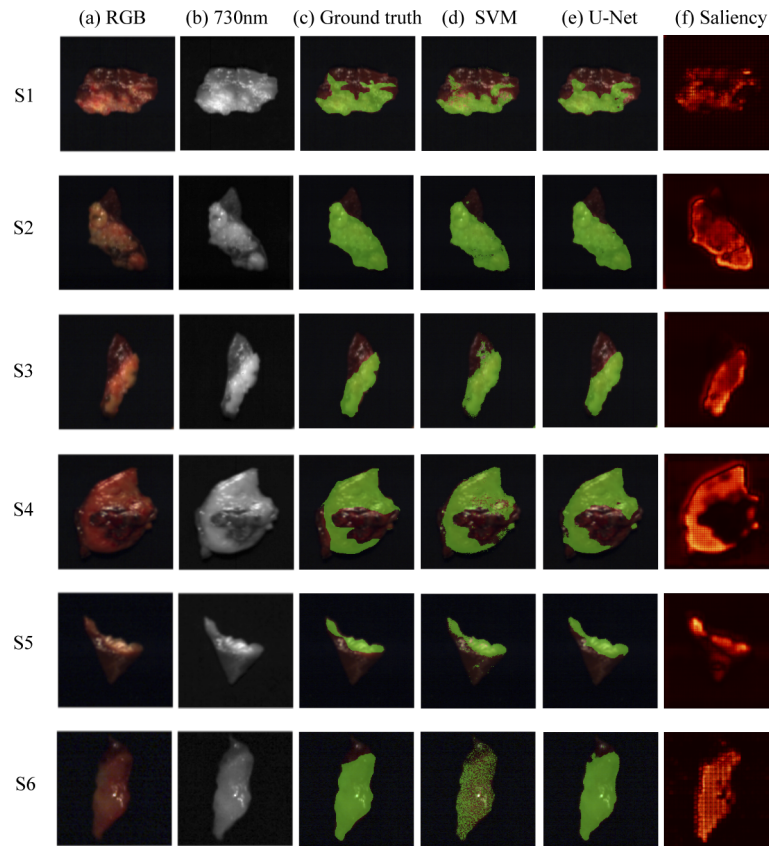


Fig. 7. Tumor delineation results (Row S1 and S6 are ICC samples, the others are HCC samples). Pixels colored with green indicate the labeled or predicted malignant area. The columns present the image of (a) the synthetic RGB images from the hyperspectral images; (b) grayscale images under 730 nm; (c) ground truth images on gross specimens; (d) overlaid results of the baseline method; (e) overlaid results of the developed method; (f) the spatial saliency map. Pixels colored with red indicate areas that are important to the model's decision, and the brighter the pixel, the more important it is.

Table 2. Ablation over model size and training parameters

EXP	Receptive field			Channel size		Patch per image			Batch size		Norm-alization	Spectral augmentation	Class reweight	Average performance			
	1	3	5	64	256	10	30	50	10	20				Acc	Sen	Spe	DSC
E0	✓				✓		✓		✓			✓		93.12	94.39	84.37	92.29
E1	✓				✓		✓		✓			✓		91.78	93.49	79.65	91.08
E2			✓		✓		✓		✓			✓		93.16	94.48	87.22	92.91
E3	✓			✓			✓		✓			✓		93.06	93.75	83.54	92.39
E4	✓				✓		✓		✓			✓		92.37	93.96	80.56	91.69
E5	✓				✓			✓	✓			✓		92.48	93.89	82.60	91.68
E6	✓				✓		✓			✓		✓		92.52	93.75	83.54	91.98
E7	✓				✓		✓		✓		✓	✓		77.51	100.00	0.00	64.99
E8	✓				✓		✓		✓					92.50	94.02	81.79	91.61
E9	✓				✓		✓		✓			✓	✓	92.31	93.33	84.41	91.94

performance, indicating the importance of combining spatial information. Second, different from previous research work [14], adding spectral normalization degraded the performance significantly in our experiments (E0 and E7). A possible explanation for this could be that the difference in spectra amplitude is quite helpful in distinguishing liver tumor tissues. The last facet is the spectral data augmentation (E8 and E0). Compared to noise-free data, the result proves that adding some small perturbations, e.g., radiation noise, improves the model's performance.

Intra-patient study We also carried out the experiment under an intra-patient framework using the same deep learning architecture as the above inter-patient settings, but the model is trained and tested on specimens from the same patient, i.e., 90% data is split as train-validation set while the rest as the test set. The intra-patient experiment reaches an accuracy of $97.91 \pm 1.94\%$, a sensitivity of $97.80 \pm 1.95\%$, a specificity of $99.28 \pm 0.91\%$ and a DSC of $98.47 \pm 1.17\%$, surpassing the inter-patient framework by a large margin.

The result indicates that spectra of tissues from the same patient have a much better correspondence than that from different patients. This would be helpful in the long-term follow-up study of patients. We could track the condition of patients before, during, and after tumor resection for quantitative analysis [12]. In addition, compared with the inter-patient study, the result confirms the fact that tumor heterogeneity exists across patients. Although the same type of tumor has common characteristics, various causes (such as alcoholic fatty liver disease, cirrhosis, and HBV infection) lead to variance in their appearance, which greatly increases the difficulty of clinical inter-patient research. Therefore, a large amount of data is needed in the future to build a database that can cover the difference between patients.

3.2.2. Interpretability of the hyperspectral image

Salient pixels As a visual result of the classification branch, the spatial saliency depicted in Fig. 7(f) indicates where the model tells tumorous tissues from non-tumorous tissues. When the tissue is judged to be cancer, the pixels that play a major role in this judgment have a greater response and appear to be brighter. Similarly, we used the DSC metric to measure the similarity between the segmentation result (Fig. 7(e)) and the saliency map (Fig. 7(f)) by comparing their binary masks of malignant tissues. The results reported a DSC of $95.80 \pm 3.29\%$, which indicates that the highlighted area in this saliency map has a high degree of consistency with the tumor segmentation results, confirming the reliability of the model. In fact, the saliency map itself is also an effective tool for segmentation. In this work, we merely use it as circumstantial evidence of the segmentation results to help us better understand and confirm the model's decision. The spatial saliency map can be regarded as another method to obtain the tumor margin. It can also be further used to improve the segmentation results in the future.

Salient wavelengths Fig. 8(a) shows the saliency scores of every spectral channel, which guides us to pay attention to the difference between normal tissue and malignant tissue from the perspective of spectral information to establish a connection with tissue optics and even clinical pathology.

The contributions of different spectral channels are depicted as the gray curve shown in Fig. 8(a). Here the spectral saliency scores are min-max normalized into the range of [0, 1]. As can be seen that not all channels contribute to this task. Regions near 420 nm, 516 nm, 607 nm, 702 nm, 767 nm, 856 nm and 975 nm reach the locally largest contributions, respectively. Through the comparison with the average spectra curve (from Fig. 5(d)), differences like slopes (such as 420 nm, 607 nm, and 856 nm) or peaks (e.g., 516 nm) can be easily observed as some intuitive features. These evidence provide some shallow explanations for the model's decision, indicating the model might use these features to distinguish different curves.

To further specify the discriminative channels, we ranked them according to their saliency scores and made a simple redundancy test. Less relevant channels are eliminated proportionally. Experiment results depicted in Fig. 8(b) reveal that nearly 90% of the spectral information is

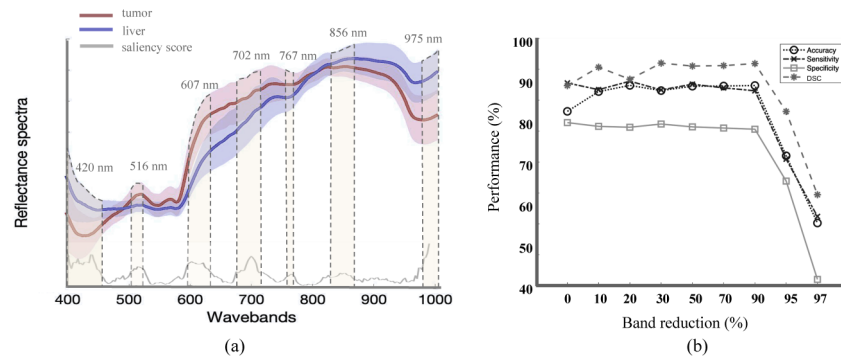


Fig. 8. Spectral saliency scores. (a) Salient channels are mapped to the average spectral curves of the malignant and normal tissues from Fig. 5; (b) Channel redundancy. Image shows the performance of the model when we proportionally reduce the number of spectra according to the ranking of the spectral saliency scores.

redundant in distinguishing malignant tissues since the top 10% salient spectral channels retain a similar performance as the 100% does. After the proportion of channels drops below 10%, the performance begins to decrease accordingly. In addition, performance improvement can be observed during channel reduction, which proves the importance of channel denoising. Although we found large redundancy exists in the spectral dimension, channel selection through saliency ranking is not optimal as we discussed in Section 2.6 that adjacent channels are most likely selected from the same region. Fewer predominant channels can possibly be found. Before that, we make an assumption that fewer spectral channels can achieve higher performances and are supposed to have a greater possibility of approaching the key factors that can distinguish tissues.

Channel selection using saliency scores To further find the dominant channels that contain discriminative information, we apply the proposed saliency-weighted channel selection to narrow the range of the salient channels. The most salient channels are selected by the saliency-weighted QRcp and the DPP model, respectively. Results are compared with that of using all channels.

To extract the dominant channels, we compared the performance of the proposed saliency-weighted methods with other channel selection strategies: (1) random selection; (2) selection according to the ranking of saliency scores; (3) manually selection according to the saliency peaks; and (4) selection using the QRcp or DPP model of spectral features without saliency scores. Empirically, we found that 5 channels could achieve a comparable performance (with roughly less than 1% drop) as 224 channels can perform. Thus, we suggest that 5 channels are used in this tumor delineation task. Specifically, the five channels extracted by saliency-weighted QRcp are 420 nm, 440 nm, 605 nm, 702 nm and 1000 nm. Slightly differently, saliency-weighted DPP extracts 420 nm, 605 nm, 702 nm, 847 nm and 1000 nm. The performance of different selection strategies can be found in Table 3. The extracted channels are pointed out in Fig. 9(a).

The result in Fig. 9 shows the importance of considering the correlation between spectral channels. Not surprisingly, random selection achieved an unsatisfactory result since noise has no contribution to this decision-making process. If we select only by ranking the salient scores, correlated channels might be redundantly selected (Fig. 9(A1)). Although we can manually decide channels that reach the local maximum saliency scores (Fig. 9(A2)), this is not the optimal way. In addition to the time-consuming disadvantages, different cross-validation settings cause subtle fluctuations in the spectral saliency map, leading to large uncertainties in manual selection. In this experiment, we applied the saliency scores as a task-specific prior and strategies of QRcp and the DPP model to measure the correlation between spectra to avoid selecting highly-correlated channels in the task. However, if we only analyze the similarity of spectra

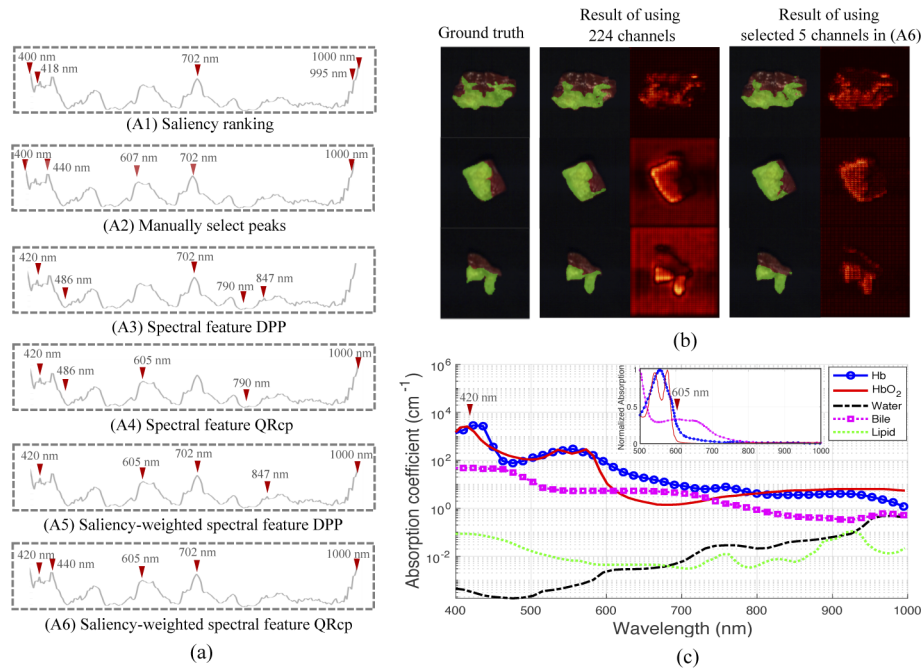


Fig. 9. Channel selection results. (a) The 5 dominant channels selected by different strategies. As can be seen that channels selected by saliency-weighted methods (A5-A6) achieve a diverse characteristic. (b) Visual inspection of the delineation results and spatial saliency map before and after channel selection. In addition to the improvement on delineation results, the salient map becomes more explicit after channel reduction, indicating the importance of discarding the redundant information in the decision. (c) Absorption coefficients of the main absorbers in the liver (Hb, HbO₂, bile, water and lipid) from 400 to 1000 nm in a logarithmic scale. The inset figure shows the normalized absorption coefficients of Hb, HbO₂, and bile between 500 and 1000 nm. Figures are adapted from [40]. Large absorption peaks of Hb and bile situated at 420 nm and 605 nm are pointed out, respectively.

Table 3. Results of different channel selection strategies (kernel size = 3)

	Accuracy	Sensitivity	Specificity	DSC
Using 224 spectra	93.06	93.75	83.54	92.39
Randomly select 5 spectra	69.94	76.04	58.18	79.20
Select by saliency ranking (no diversity)	81.71	81.43	77.76	87.56
Manually select top spectra	91.35	92.58	82.97	93.89
Select by DPP	91.15	85.49	85.92	93.65
Select by saliency-weighted DPP	93.08	93.20	84.95	95.11
Select by QRcp	91.08	90.95	85.50	92.45
Select by saliency-weighted QRcp	91.90	91.78	86.68	93.04

without considering saliency, it is difficult for the model to select the optimal subset to achieve this task. As shown in Fig. 9(A3-A4), channels located at the valley of saliency scores were also included in the subset (e.g., 486 nm, 790 nm), which apparently does not apply to this task, because their contribution is not significant. Therefore, we performed the saliency-weighted channel selection to automatically find the dominant channels. Compared in Table 3, the two strategies achieved superior performances than others, reporting the same dominant channels to discriminate primary liver tumor at 420 nm, 605 nm, 702 nm and 993 nm with a different selection at 440 nm and 847 nm (Fig. 9(A5-A6)). The DPP is a randomized method for selecting subsets, while QRcp is a deterministic method for selecting orthogonal features. In practical use, both strategies can well extract the key channels for specific tasks. In practice, the QR strategy is preferable due to its deterministic nature.

The results also suggest that extracting a small but discriminative subset can effectively reduce data noise and even achieve better performance than using all spectra. As can be seen that the highlighted irrelevant pixels, such as the background, in the spatial saliency map have been reduced after channel pruning (see Fig. 9(b)), indicating a higher reliability in the model decision-making process.

Possible indications From the perspective of tissue optics, we consider several main absorbers in the liver, i.e., oxygenated and deoxygenated hemoglobin (HbO_2 and Hb), water, lipid, and bile, some qualitative speculations can be obtained.

Normal liver tissue typically receives its blood supply from hepatic portal vein, while primary liver tumor tissue receives most of the blood supply from hepatic arterial. Thus, the hemoglobin absorption bands could be very different. Specifically, Hb has the highest absorption peak at 420 nm while HbO_2 at 410 nm [8]. The result of salient channels at 420 nm is closely related to the absorption peaks of Hb and HbO_2 . Another important indicator in the visible is bile. Healthy liver tissue allows the perfusion of bile, while liver tumors, such as ICC, have different structures, resulting in different bile perfusions. The result of 605 nm channel obtained through channel selection is in close agreement with the previous work [40] where they report bile has a large absorption peak centered exactly at 605 nm (Fig. 9(c)). As the wavelength increases in the near-infrared region, water absorption prevails, and tends to be consistent with bile due to the fact that bile is mainly composed of water [40]. The spectral signature around 1000 nm provides possible evidence of differences in water content of bile or blood between the two tissues. In addition, scattering becomes more pronounced than absorption in the near-infrared band. Therefore, we suggest that the significant channels at 702nm, 847nm, and 1000nm need to be combined with the scattering characteristics for future analysis, such as changes in the nucleus and organelles during tissue development that cause changes in the light scattering characteristics.

4. Discussion

HSI is a promising technology for future use in surgical applications with non-contact, label-free and fast imaging characteristics. In this study, we evaluated the effectiveness of the application of HSI and deep learning framework on fresh liver specimens for tumor delineation. Since a human-intelligible result is more acceptable than a result from an undetectable black box, especially for practical use in the medical domain, we emphasize the importance of adding explanations to the model's decision. Specifically, a saliency map-based analysis is performed on medical hyperspectral images from both the spatial and spectral aspects. On the basis of spectral saliency, we proposed a novel method to further extract key channels with possible indications and also select filters for a multi-spectral camera that would be used practically in the future. To the best of our knowledge, this is the first work studying intra-operative liver specimens using HSI. We focus on primary liver tumors, but the methods proposed in this work can also be applied to other tumor types.

In terms of automatic tumor delineation, the developed U-Net framework outperforms the pixel-wise classification method by a large margin and the ablation study shows that larger receptive fields produce better performance. These results indicate the importance of using spatial features during the learning process. In addition, we added a classification head to the network architecture as the second output of the image-level classification result. In this work, we merely use it as the explanation of segmentation results to confirm the model's decision is trustworthy. In subsequent explorations, this multi-task framework could also be extended to perform pathological classification, e.g., to distinguish HCC from ICC and identify the meaningful wavelengths which are responsible for distinguishing subtle pathological differences. In practice, the margin of ICC is more difficult to identify and the curative resection rate of ICC is lower than that of HCC [41,42]. Therefore, more research can be carried out on the identification of ICC margin. Limited by the amount of data, we did not conduct further pathological analysis in this experiment. Currently, we are collecting more specimens for further pathological exploration. In the effort to explore the model's generalization capability across different patients, the results reveal an obvious inter-individual variance compared with the intra-patient performance, which reinforces further the fact that tumor heterogeneity is a non-negligible factor in the clinical study. In that sense, again a larger data amount is needed to cover more cases.

As regards to the result of spectral interpretability, we demonstrate that the benefits of exploring salient spectra are two-fold: First, significant dimension reduction explicitly shows the contributions of different spectral channels, thus revealing an understandable decision making process of the model. In the result, we have reached a conclusion that Hb and bile content may be indicative in distinguishing liver tumors from normal tissues. Similarly, we can also find meaningful spectral features from the metabolic differences between tumors and normal tissues. These two different ideas can take advantage of each other's insight to better understand the process of tumor development. For example, previous studies reported possible association of bile acids and cancer development [43]. The bile acid profile in patients with primary liver cancer is significantly different with that in healthy people. The alterations in bile acids may affect hepatic metabolic homeostasis and contribute to the pathogenesis and development of liver cancer. [44–46]. We believe it would be meaningful to further explore the spectral characteristics of bile acids. However, the wavebands used in this experiment are in 400–1000 nm, and the results that can be obtained are limited to chromophores that are predominant in this waveband. Follow-up explorations can also consider establishing more possible connections from the characteristic bands of tumor metabolism, such as glucose and lipids, etc [47–51], at corresponding wavebands.

The second is that reducing the spectral dimension is beneficial to hardware design in that it can effectively help filter selection for a specific application scenario and also reduce data collection time and model complexity. The hardware design of a multi-spectral imaging system is a necessary step towards future intra-operative study. Back to the motivation of this study, our purpose is to provide surgeons with direct feedback of pathological indications during the operation prior to the time-consuming post-operative pathology report. This requires HSI settings to measure the tissues in real-time. As far as we know, all previous studies, including our preliminary efforts, are based on ex-vivo specimens or using line-scanning hyperspectral cameras as the first step towards clinical use in the operation. It is currently almost impossible to actually use them on the surgical bed. According to the principle of HSI system design [52], there is always a trade-off among data acquisition time, spatial resolution, and spectral resolution since a real-time acquisition of three-dimensional data by a two-dimensional image sensor is a challenge in itself. We always have to sacrifice some of the performance of the system to meet the requirements of specific application scenarios, that is, a task-specific design is required. In the case of a real-time medical application, line-scanning or spectrum-scanning HSI systems are not suitable because they need a long data acquisition time. We have to ensure fast imaging speed and a large field of view for careful inspection, so a sacrifice on spectral channels is inevitable.

Fortunately, studies proved that only a small subset of spectra is sufficient to reproduce or even improve the best performance of using all channels. This is easy to understand because hundreds of spectral channels not only bring abundant information during the training process but also bring up a lot of noise. Therefore, the saliency-weighted channel pruning method provides a novel way to select the band of the multi-spectral imaging system for future spectral imaging applications in real-time intraoperative tasks.

5. Conclusions

In conclusion, this work evaluated the feasibility of HSI for tumor delineation on fresh ex-vivo liver specimens using a deep learning framework. In this work, we developed a U-Net segmentation framework and added a saliency map-based post-hoc explanation on the hyperspectral images to explore the spatial and spectral interpretability of the model's decision. Indicatively, the selected spectral channels are closely related to the differences of average spectral curves and the absorption coefficients between the normal and malignant tissues. Those salient spectral channels can be further pruned to form a relatively small subset to perform a real-life task without degrading the performance. This work preliminarily confirmed the feasibility and effectiveness of HSI in delineating liver tumors. Future work will focus on improving HSI performance required for intra-operative use, paving the way for real-time assistance during surgery.

Funding. Beijing National Research Center For Information Science And Technology (BNR2019ZS01005); CAMS Innovation Fund for Medical Sciences (CIFMS) (2020-I2M-C&T-B-019); WBE Liver Fibrosis Foundation (CFHPC 2020021); Simons Foundation (504054).

Acknowledgments. XZ is supported by the Simmons Foundation (504054).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer J. Clinicians* **68**(6), 394–424 (2018).
2. A. W. Chan, J. Zhong, S. Berhane, H. Toyoda, A. Cucchetti, K. Shi, T. Tada, C. C. Chong, B.-D. Xiang, L.-Q. Li, P. B. S. Lai, V. Mazzaferro, M. García-Fiñana, M. Kudo, K. Takashi, S. Roayaie, and P. J. Johnson, "Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection," *J. Hepatol.* **69**(6), 1284–1293 (2018).
3. M. Garancini, L. Gianotti, F. Romano, V. Giardini, F. Uggeri, and G. Torzilli, "The role of ultrasound in hepatic surgery," *Hepatic Surg* (2013).
4. Y.-M. Zhang, R. Shi, J.-C. Hou, Z.-R. Liu, Z.-L. Cui, Y. Li, D. Wu, Y. Shi, and Z.-Y. Shen, "Liver tumor boundaries identified intraoperatively using real-time indocyanine green fluorescence imaging," *J. Cancer Res. Clin. Oncol.* **143**(1), 51–58 (2017).
5. Y. Nakaseko, T. Ishizawa, and A. Saiura, "Fluorescence-guided surgery for liver tumors," *J. Surg. Oncol.* **118**(2), 324–331 (2018).
6. B. Fei, "Hyperspectral imaging in medical applications," in *Data Handling in Science and Technology*, vol. 32 (Elsevier, 2020), pp. 523–565.
7. G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *J. Biomed. Opt.* **19**(1), 010901 (2014).
8. S. L. Jacques, "Optical properties of biological tissues: a review," *Phys. Med. Biol.* **58**(11), R37–R61 (2013).
9. H. Fabelo, S. Ortega, A. Szolna, D. Bulters, J. F. Piñeiro, S. Kabwama, A. JO'Shanahan, H. Bulstrode, S. Bisshopp, B. R. Kiran, D. Ravi, R. Lazcano, D. Madroñal, C. Sosa, E. Carlos, M. Marquez, M. D. L. L. Plaza, R. Camacho, D. Carrera, M. Hernández, G. M. Callicó, J. M. Molina, B. Stanciulescu, G.-Z. Yang, R. S. Juárez, C. Sanz, and R. Sarmiento, "In-vivo hyperspectral human brain image database for brain cancer detection," *IEEE Access* **7**, 39098–39116 (2019).
10. H. Fabelo, S. Ortega, R. Lazcano, D. Madroñal, G. M. Callicó, E. Juárez, R. Salvador, D. Bulters, H. Bulstrode, A. Szolna, J. F. Piñeiro, C. Sosa, A. J. O'Shanahan, S. Bisshopp, M. Hernández, J. Morera, D. Ravi, B. R. Kiran, A. Vega, A. Bez-Quevedo, G.-Z. Yang, B. Stanciulescu, and R. Sarmiento, "An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation," *Sensors* **18**(2), 430 (2018).

11. F. Manni, F. van der Sommen, H. Fabelo, S. Zinger, C. Shan, E. Edström, A. Elmi-Terander, S. Ortega, G. Marrero Callicó, and P. H. N. d. With, "Hyperspectral imaging for glioblastoma surgery: Improving tumor identification using a deep spectral-spatial approach," *Sensors* **20**(23), 6955 (2020).
12. G. Lu, J. V. Little, X. Wang, H. Zhang, M. R. Patel, C. C. Griffith, M. W. El-Deiry, A. Y. Chen, and B. Fei, "Detection of head and neck cancer in surgical specimens using quantitative hyperspectral imaging," *Clin. Cancer Res.* **23**(18), 5426–5436 (2017).
13. B. Fei, G. Lu, X. Wang, H. Zhang, J. V. Little, M. R. Patel, C. C. Griffith, M. W. El-Deiry, and A. Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *J. Biomed. Opt.* **22**(08), 1 (2017).
14. E. Kho, L. L. De Boer, K. K. Van de Vijver, F. Van Duijnhoven, M.-J. T. V. Peeters, H. J. Sterenberg, and T. J. Ruers, "Hyperspectral imaging for resection margin assessment during cancer surgery," *Clin. Cancer Res.* **25**(12), 3572–3580 (2019).
15. E. Kho, B. Dashbozorg, L. L. De Boer, K. K. Van de Vijver, H. J. Sterenberg, and T. J. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," *Biomed. Opt. Express* **10**(9), 4496–4515 (2019).
16. M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, and B. Fei, "Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning," *Biomed. Opt. Express* **11**(3), 1383–1400 (2020).
17. J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inf. Decis. Making* **20**(1), 310 (2020).
18. H. Lee, S. Yune, M. Mansouri, M. Kim, S. H. Tajmir, C. E. Guerrier, S. A. Ebert, S. R. Pomerantz, J. M. Romero, S. Kamalian, R. G. Gonzalez, M. H. Lev, and S. Do, "An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets," *Nat. Biomed. Eng.* **3**(3), 173–182 (2019).
19. K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," arXiv preprint arXiv:1312.6034 (2013).
20. A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of Imaging* **6**(6), 52 (2020).
21. L. L. de Boer, E. Kho, J. Nijkamp, K. K. Van de Vijver, H. J. Sterenberg, L. C. ter Beek, and T. J. Ruers, "Method for coregistration of optical measurements of breast tissue with histopathology: the importance of accounting for tissue deformations," *J. Biomed. Opt.* **24**(07), 1 (2019).
22. M. Halicek, J. V. Little, X. Wang, Z. G. Chen, M. Patel, C. C. Griffith, M. W. El-Deiry, N. F. Saba, A. Y. Chen, and B. Fei, "Deformable registration of histological cancer margins to gross hyperspectral images using demons," in *Medical Imaging 2018: Digital Pathology*, vol. 10581 (International Society for Optics and Photonics, 2018), p. 105810N.
23. G. Lu, L. Halig, D. Wang, Z. G. Chen, and B. Fei, "Hyperspectral imaging for cancer surgical margin delineation: registration of hyperspectral and histological images," in *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9036 (International Society for Optics and Photonics, 2014), p. 90360S.
24. D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE Trans. Med. Imaging* **18**(8), 712–721 (1999).
25. G. Andrew and J. Gao, "Scalable training of l1-regularized log-linear models," in *Proceedings of the 24th international conference on Machine learning*, (2007), pp. 33–40.
26. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
27. P. Hensman and D. Masko, "The impact of imbalanced training data for convolutional neural networks," Degree Project in Computer Science, KTH Royal Institute of Technology (2015).
28. W. Li, C. Chen, M. Zhang, H. Li, and Q. Du, "Data augmentation for hyperspectral image classification with deep cnn," *IEEE Geosci. Remote Sensing Lett.* **16**(4), 593–597 (2019).
29. W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.* **7**(2), 118–139 (2019).
30. S. J. Wirkert, F. Isensee, A. S. Vemuri, K. Maier-Hein, B. Fei, and L. Maier-Hein, "Domain and task specific multispectral band selection (conference presentation)," in *Design and Quality for Biomedical Technologies XI*, vol. 10486 (International Society for Optics and Photonics, 2018), p. 104860H.
31. S. J. Wirkert, N. T. Clancy, D. Stoyanov, S. Arya, G. B. Hanna, H.-P. Schlemmer, P. Sauer, D. S. Elson, and L. Maier-Hein, "Endoscopic sheffield index for unsupervised in vivo spectral band selection," in *International Workshop on Computer-Assisted and Robotic Endoscopy* (Springer, 2014), pp. 110–120.
32. D. Nouri, Y. Lucas, and S. Treuillet, "Hyperspectral interventional imaging for enhanced tissue visualization and discrimination combining band selection methods," *Int. J. CARS* **11**(12), 2185–2197 (2016).
33. V. Sharma, A. Diba, T. Tuytelaars, and L. Van Gool, "Hyperspectral cnn for image classification & band selection, with application to face recognition," Technical report KUL/ESAT/PSI/1604, KU Leuven, ESAT, Leuven, Belgium (2016).
34. J. N. Kutz, *Data-driven modeling & scientific computation: methods for complex systems & big data* (Oxford University Press, 2013).
35. M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing qr factorization," *SIAM J. Sci. Comput.* **17**(4), 848–869 (1996).

36. J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, “Determinantal processes and independence,” *Probab. Surveys* **3**, 206–229 (2006).
37. A. Kulesza and B. Taskar, “Determinantal point processes for machine learning,” arXiv preprint arXiv:1207.6083 (2012).
38. S. Chakroborty and G. Saha, “Feature selection using singular value decomposition and qr factorization with column pivoting for text-independent speaker identification,” *Speech Commun.* **52**(9), 693–709 (2010).
39. G. Anthony, H. Gregg, and M. Tshilidzi, “Image classification using svms: one-against-one vs one-against-all,” arXiv preprint arXiv:0711.2914 (2007).
40. R. Nachabé, D. J. Evers, B. H. Hendriks, G. W. Lucassen, M. van der Voort, J. Wesseling, and T. J. Ruers, “Effect of bile absorption coefficients on the estimation of liver tissue optical properties and related implications in discriminating healthy and tumorous samples,” *Biomed. Opt. Express* **2**(3), 600–614 (2011).
41. K. Lafaro, M. S. Grandhi, J. M. Herman, and T. M. Pawlik, “The importance of surgical margins in primary malignancies of the liver,” *J. Surg. Oncol.* **113**(3), 296–303 (2016).
42. X.-D. Zhou, Z.-Y. Tang, J. Fan, J. Zhou, Z.-Q. Wu, L.-X. Qin, Z.-C. Ma, H.-C. Sun, S.-J. Qiu, Y. Yu, N. Ren, Q.-H. Ye, L. Wang, and S.-L. Ye, “Intrahepatic cholangiocarcinoma: report of 272 patients compared with 5, 829 patients with hepatocellular carcinoma,” *J. Cancer Res. Clin. Oncol.* **135**(8), 1073–1080 (2009).
43. A. Jusakul, N. Khuntikeo, W. G. Haigh, R. Kuver, G. N. Ioannou, W. Loilome, N. Namwat, V. Bhudhisawasdi, A. Pugkhem, C. Pairojkul, and P. Yongvanit, “Identification of biliary bile acids in patients with benign biliary diseases, hepatocellular carcinoma and cholangiocarcinoma,” *Asian Pac J Cancer Prev* **13**, 77–82 (2012).
44. D. O. Rees, P. J. Crick, G. J. Jenkins, Y. Wang, W. J. Griffiths, T. H. Brown, and B. Al-Sarireh, “Comparison of the composition of bile acids in bile of patients with adenocarcinoma of the pancreas and benign disease,” *J. Steroid Biochem. Mol. Biol.* **174**, 290–295 (2017).
45. E. Danese, D. Negrini, M. Pucci, S. De Nitto, D. Ambrogi, S. Donzelli, P. M.-J. Lievens, G. L. Salvagno, and G. Lippi, “Bile acids quantification by liquid chromatography–tandem mass spectrometry: Method validation, reference range, and interference study,” *Diagnostics* **10**(7), 462 (2020).
46. X. Zhang, Z. Yang, Z. Shi, Z. Zhu, C. Li, Z. Du, Y. Zhang, Z. Wang, Z. Jiao, X. Tian, J. Zhang, W. Zhai, and Q. Kan, “Analysis of bile acid profile in plasma to differentiate cholangiocarcinoma from benign biliary diseases and healthy controls,” *J. Steroid Biochem. Mol. Biol.* **205**, 105775 (2021).
47. M. Mossenta, D. Busato, M. Dal Bo, and G. Toffoli, “Glucose metabolism and oxidative stress in hepatocellular carcinoma: role and possible implications in novel therapeutic strategies,” *Cancers* **12**(6), 1668 (2020).
48. S. De Matteis, A. Ragusa, G. Marisi, S. De Domenico, A. Casadei Gardini, M. Bonafè, and A. M. Giudetti, “Aberrant metabolism in hepatocellular carcinoma provides diagnostic and therapeutic opportunities,” *Oxid. Med. Cell. Longevity* **2018**, 1–13 (2018).
49. E. D. Pope III, E. O. Kimbrough, L. P. Vemireddy, P. K. Surapaneni, J. A. Copland III, and K. Mody, “Aberrant lipid metabolism as a therapeutic target in liver cancer,” *Expert Opin. Ther. Targets* **23**(6), 473–483 (2019).
50. E. B. Evangelista, S. A. Kwee, M. M. Sato, L. Wang, C. Rettenmeier, G. Xie, W. Jia, and L. L. Wong, “Phospholipids are a potentially important source of tissue biomarkers for hepatocellular carcinoma: Results of a pilot study involving targeted metabolomics,” *Diagnostics* **9**(4), 167 (2019).
51. P. Shao, D. Sun, L. Wang, R. Fan, and Z. Gao, “Deep sequencing and comprehensive expression analysis identifies several molecules potentially related to human poorly differentiated hepatocellular carcinoma,” *FEBS Open Bio.* **7**(11), 1696–1706 (2017).
52. Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, “Review of spectral imaging technology in biomedical engineering: achievements and challenges,” *J. Biomed. Opt.* **18**(10), 100901 (2013).